

AD-A186 036

TEST OF LINEARITY IN GENERAL REGRESSION MODELS(U)
PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS
X R CHEN ET AL. DEC 86 TR-86-49 AFOSR-TR-87-1093

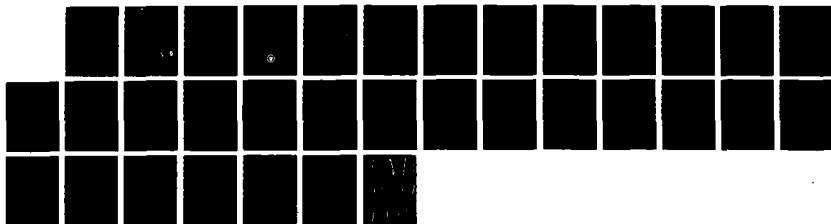
1/1

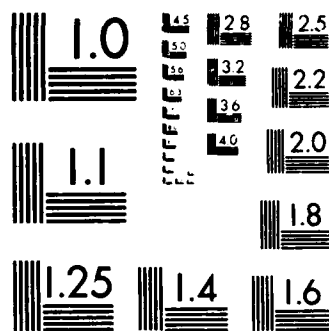
UNCLASSIFIED

F49620-85-C-0008

F/G 12/3

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

DTIC FILE COPY

Unclassified

1

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

AFOSR-TR-87-1093

4. TITLE (and Subtitle)

Test of linearity in general regression models

5. TYPE OF REPORT & PERIOD COVERED

Technical - December 1986

Journal

6. PERFORMING ORG. REPORT NUMBER

86-49

7. AUTHOR(s)

X.R. Chen and P.R. Krishnaiah

8. CONTRACT OR GRANT NUMBER(s)

F49620-85-C-0008

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Center for Multivariate Analysis

Fifth Floor Thackeray Hall

University of Pittsburgh, Pittsburgh, PA 15260

10. PROGRAM ELEMENT, PROJECT, TASK
AREA & WORK UNIT NUMBERS

611021-

2304 AS

11. CONTROLLING OFFICE NAME AND ADDRESS

Air Force Office of Scientific Research

Department of the Air Force

Bolling Air Force Base, DC 20332

(111)

12. REPORT DATE

December 1986

13. NUMBER OF PAGES

29

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

SCME 98-11

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING
SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear regression, consistency, strong approximation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let (\bar{X}_i, Y_i) , $i = 1, \dots, n$, be iid. samples of (\bar{X}, Y) . This paper proposes a method for testing the linearity of the regression function $E(Y|\bar{X} = x)$. The asymptotic distribution (under null hypothesis) and the

DTIC
ELECTE
OCT 07 1987
S E D

AD-A186 036

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20

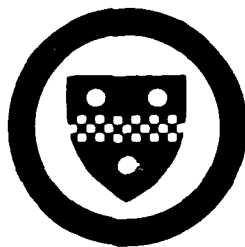
asymptotic power of the test are determined. Also, consistency of the test is proved under mild conditions.

AFOSR-TR. 87-1093

TEST OF LINEARITY IN
GENERAL REGRESSION MODELS *

X.R. Chen and P.R. Krishnaiah
Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis
University of Pittsburgh



87 9 24 281

TEST OF LINEARITY IN
GENERAL REGRESSION MODELS *

X.R. Chen and P.R. Krishnaiah
Center for Multivariate Analysis
University of Pittsburgh

December 1986

Technical Report No. 86-49

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input checked="" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

- * Research sponsored by the Air Force Office of Scientific Research (AFSC), under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

TEST OF LINEARITY IN GENERAL REGRESSION MODELS *

X.R. Chen and P.R. Krishnaiah

Center for Multivariate Analysis
University of Pittsburgh

ABSTRACT

Let (\bar{X}_i, Y_i) , $i = 1, \dots, n$, be iid. samples of (\bar{X}, Y) . This paper proposes a method for testing the linearity of the regression function $E(Y|\bar{X} = x)$. The asymptotic distribution (under null hypothesis) and the asymptotic power of the test are determined. Also, consistency of the test is proved under mild conditions.

AMS 1980 Subject Classification: Primary 62J05; Secondary 62C35.

Key words and phrases: Linear regression, consistency, strong approximation.

* Research sponsored by the Air Force Office of Scientific Research (AFSC), under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

1. INTRODUCTION

Linear regression models are widely used in statistical analysis of experimental and observational data. Usually the linearity of the model is merely an assumption and cannot be taken for granted. In some planned experiments, repeated measurements on the dependent variable Y can be taken while the independent variable X is held fixed. In such cases standard analysis-of-variance technique can be employed to generate a test for linearity. In many applications, however, the independent variable is observed simultaneously with Y . That is to say, X , as well as Y , is a random variable. Under such circumstance the usual method for testing linearity cannot apply.

In this paper we shall study this problem in the large-sample context. We propose a method to test the linearity hypothesis based on a grouping of the data. The critical value of test-statistic is determined so that the test has a prescribed level of significance α asymptotically as the sample size tends to infinity. The consistency of the test is established, and the asymptotic power is calculated when the distance (in some sense) between the true regression function and the space of linear functions tends to zero in some specific rate.

2. ASSUMPTIONS OF THE MODEL

In this section we shall give a detailed account of our assumptions. We shall adhere to these assumptions in the sequel.

A basic assumption is that the conditional distribution of Y given $\bar{X} = x$ has a form $F(y - m(x))$, i.e., $P(Y < y | \bar{X} = x) = F(y - m(x))$, where F is a fixed distribution function. We do not assume F to be known. Let $(\bar{X}_1, Y_1), \dots, (\bar{X}_n, Y_n)$ be iid. observations of (\bar{X}, Y) . Then, under this assumption, we can give Y_i a more convenient expression, as follows:

$$Y_i = m(\bar{X}_i) + e_i, \quad i = 1, \dots, n \quad (1)$$

where e_1, e_2, \dots are iid. with common distribution function F , and $\bar{X}_1, \dots, \bar{X}_n, e_1, \dots, e_n$ are mutually independent. Also, $Ee_1 = 0$. We further assume that

1°. The moment generating function of e_1 exists in some neighborhood of zero.

2°. The variance σ^2 of e_1 (whose existence follows from 1°) is positive. Denote by μ the probability distribution of \bar{X} . We assume that

3°. $E\|\bar{X}\|^2 < \infty$, and $\text{COV}(\bar{X}) > 0$ ($\|\cdot\|$ is the Eculidian norm).

4°. μ has no singular component, and if μ has an absolute component with density f , then for sufficiently small $a > 0$ there exists an open set G_a , such that

$$|G_a \Delta \{x: f(x) > a\}| = 0$$

where $|A|$ denotes the Lebesgue measure of A , and Δ is symmetric difference. This condition is a rather weak one since it is satisfied by such f whose discontinuity points all lie in a closed set with Lebesgue measure zero.

For brevity, in the sequel we shall use "model (1)" in the sense that all of the above assumptions are met.

3. CONSTRUCTION OF THE TEST

Choose two numbers ε_1 and ε_2 , with the condition

$$0 < \varepsilon_1 < \varepsilon_2 < 1/3. \quad (2)$$

Some further restrictions on the choice of $\varepsilon_1, \varepsilon_2$ will be needed, which will be stated in Section 5.

Put $\ell_n = n^{-(1/3-\varepsilon_2)/d}$. Decompose R^d into a set J_n^* of supercube having the form

$$\{(x^{(1)}, \dots, x^{(d)}): a_i \ell_n \leq x^{(i)} < (a_i + 1) \ell_n, \quad i = 1, \dots, d\}, \quad (3)$$

$$a_i = 0, \pm 1, \pm 2, \quad i = 1, \dots, d.$$

For $J \in J_n^*$, use $\#(J)$ to denote the number of elements in the set $J \cap \{\bar{X}_1, \dots, \bar{X}_n\}$. Write

$$\{J: J \in J_n^*, \#(J) \geq n^{2/3+\varepsilon_1}\} = \{J_{n1}, J_{n2}, \dots, J_{nc_n}\}. \quad (4)$$

Obviously,

$$c_n \leq n^{1/3-\varepsilon_1}. \quad (5)$$

Further, let

$$J_{ni} \cap \{\bar{X}_1, \dots, \bar{X}_n\} = \{\bar{X}_{ni}(1), \bar{X}_{ni}(2), \dots, \bar{X}_{ni}(n_i)\}$$

We have by definition

$$n_i \geq n^{(2/3)+\varepsilon_1}, \quad i = 1, \dots, c_n. \quad (6)$$

We shall write $y_{ni}(j)$ and $e_{ni}(j)$ for y_k and e_k , when $\bar{X}_{ni}(j) = \bar{X}_k$.

Denote by \bar{X}_{ni} the arithmetic mean of $\bar{X}_{ni}(j)$, $j = 1, \dots, n_i$. Similarly

we define Y_{ni} and e_{ni} .

If the null hypothesis H_0 :

$$m(x) = \alpha + \beta'x \quad (7)$$

is true, we have

$$Y_{ni} = \alpha + \bar{X}_{ni}'\beta + e_{ni}, \quad i = 1, \dots, c_n \quad (8)$$

where $Ee_{ni} = 0$. $\text{Var } e_{ni} = \sigma^2/n_i$, and e_{n1}, \dots, e_{nc_n} are independent.

Applying weighted least squares method to this model with weight matrix

$$W_{(n)} = \text{diag}(n_1, n_2, \dots, n_{c_n})$$

one obtains the weighted residual sum of squares as

$$RSS_n = Y_{(n)}'W_{(n)}Y_{(n)} - Y_{(n)}'W_{(n)}\bar{X}_{(n)}\left(\bar{X}_{(n)}'W_{(n)}\bar{X}_{(n)}\right)^{-1}\bar{X}_{(n)}'W_{(n)}Y_{(n)} \quad (9)$$

where

$$\bar{X}_{(n)} = (\bar{X}_{n1}, \dots, \bar{X}_{nc_n})', \quad Y_{(n)} = (\tilde{Y}_{n1}, \dots, \tilde{Y}_{nc_n})', \quad e_{(n)} = (e_{n1}, \dots, e_{nc_n})'$$

$$\tilde{X}_{ni} = \bar{X}_{ni} - \bar{X}_n, \quad \tilde{Y}_{ni} = Y_{ni} - \bar{Y}_n, \quad \tilde{e}_{ni} = e_{ni} - \bar{e}_n, \quad i = 1, \dots, c_n$$

$$\bar{X}_n = \sum_{i=1}^{c_n} n_i \bar{X}_{ni} / N_n, \quad \bar{Y}_n = \sum_{i=1}^{c_n} n_i Y_{ni} / N_n, \quad \bar{e}_n = \sum_{i=1}^{c_n} n_i e_{ni} / N_n$$

and

$$N_n = c_1 + n_2 + \dots + n_{c_n}. \quad (10)$$

The definition of RSS_n is meaningful even when (7) is false. In case that (7) is true, we have

$$RSS_{(n)} = e_{(n)}'W_{(n)}e_{(n)} - e_{(n)}'W_{(n)}\bar{X}_{(n)}\left(\bar{X}_{(n)}'W_{(n)}\bar{X}_{(n)}\right)^{-1}\bar{X}_{(n)}'W_{(n)}e_{(n)}. \quad (11)$$

RSS_n tends to be large when the null hypothesis H_0 is not true, and this suggests a test for H_0 . Reject H_0 when

$$RSS_n / \sigma_n^2 > C \quad (12)$$

and accept H_0 when (12) does not hold. σ_n^2 is a suitable estimate of σ^2 , to be defined in Section 5. In order to determine C in (12), we must study the asymptotic distribution of the statistic RSS_n / σ_n^2 under H_0 . This will be done in the following two sections.

4. ASYMPTOTIC DISTRIBUTION OF THE RESIDUAL SUM OF SQUARES

The purpose of this section is to prove the following:

THEOREM 1. Let $P^* = P^*(\bar{X}_1, \bar{X}_2, \dots)$ be the conditional distribution given $\{\bar{X}_1, \bar{X}_2, \dots\}$. Under model (1) and when H_0 is true, there exists random variable ξ_n whose conditional distribution given $\{\bar{X}_1, \bar{X}_2, \dots\}$ is $\chi^2_{c_n-d}$, such that with probability one we have

$$RSS/\sigma^2 - \xi_n \xrightarrow{P^*} 0, \quad \text{as } n \rightarrow \infty. \quad (13)$$

Proof: First we proceed to show that

$$\bar{X}'_{(n)} W_{(n)} \bar{X}_{(n)} / n \rightarrow \text{COV}(\bar{X}), \quad \text{a.s.}, \quad \text{as } n \rightarrow \infty. \quad (14)$$

For this purpose, denote the atoms of \bar{X} by a_1, a_2, \dots , and put $p_i = P(\bar{X} = a_i) > 0, i = 1, 2, \dots$. Given $\eta > 0$, find N so that $p_{N+1} + p_{N+2} + \dots < \eta$. Find $b > 0$ small enough, such that

$$\int_{\{x=f(x) \leq b\}} f dx < \eta. \quad (15)$$

According to condition 4°, there exist an open set G , such that

$$|G \Delta \{x: f(x) > b\}| = 0.$$

Choose a bounded closed set $F_0 \subset G$, so that $\int_{G-F_0} f dx < \eta$. Denote by ϵ'_1 the distance between F_0 and the boundary of G , then $\epsilon'_1 > 0$.

By the inequality of Bennett (1962), we have

$$P(|\#(J)/n - \mu(J)| \geq \epsilon) \leq 2\exp(-n\epsilon^2/(2 + \epsilon)) \quad (16)$$

$$P(|\#(J)/n - \mu(J)| \geq \epsilon\mu(J)) \leq 2\exp(-cn\mu(J)) \quad (17)$$

where $\epsilon > 0$, and $c > 0$ is a constant depending solely on ϵ . Define two

subsets of J_n^*

$$H_1 = \{J: J \in J_n^*, a_i \in J \text{ for some } i \leq N\}$$

$$H_2 = \{J: J \in J_n^*, J \cap F_0 \neq \emptyset\}$$

and an event

$$E_n = \{\#(J) \geq n^{2/3+\epsilon_1}, \text{ for every } J \in H_1 \cup H_2\}.$$

Since the number of J 's contained in $H_1 \cup H_2$ is of the order $O(n^{1/3-\epsilon_2})$ and $\mu(J) \geq bn^{-1/3+\epsilon_1}$ for $J \in H_1 \cap H_2$ and n sufficiently large, it is easy to prove by using (16) that

$$P(E_n) \geq 1 - O(e^{-cn^{1/3}}) \quad (18)$$

where c does not depend on n .

Now define the event

$$F_n = \{\min\{\mu(I_{ni}): i = 1, \dots, c_n\} \leq n^{-1/3+\epsilon_1/2}\}$$

We are going to prove that for any given positive integer k , it is true that

$$P(F_n) = O(n^{-k}). \quad (19)$$

For this purpose, define $D_n = \{(x^{(1)}_i, \dots, x^{(d)}_i): |x^{(i)}_i| \leq n^{(k+1)/2}, i = 1, \dots, d\}$. By condition 4°, it is readily shown

$$P(\bar{X}_i \in D_n, i = 1, \dots, n) \geq 1 - O(n^{-k}).$$

Define

$$H_3 = \{J: J \in J_n^*, J \cap D_n \neq \emptyset, \mu(J) \leq n^{-1/3+\epsilon_1/2}\}.$$

There are at most $n^{(k+1)d}$ elements in H_3 . Using (16), for n sufficiently large, we have

$$\begin{aligned} P(F_n) &\leq Cn^{-k} + \sum_{J \in H_3} P(\#(J) \geq n^{2/3+\epsilon_1}) \\ &\leq Cn^{-k} + \sum_{J \in H_3} P(|\#(J)/n - \mu(J)| \geq n^{-1/3+\epsilon_1/2}) \\ &\leq Cn^{-k} + n^{(k+1)d} O(e^{-n^{1/3}}) = O(n^{-k}) \end{aligned} \quad (20)$$

which proves (19)

For each $t \in R^d$, denote by $t^{(u)}$ the u -th coordinate of t . Write $EX^{(u)} = m_u$. For each $J \in J_n^*$, choose a point $t_J \in J$, and if J is one of I_{n1}, \dots, I_{nc_n} , then choose $t_J = \bar{X}_{ni}$. Put $\tilde{J}_n = J_n^* - \{I_{ni} : i = 1, \dots, c_n\}$, we have

$$\begin{aligned} \left| \sum_{i=1}^{c_n} \frac{n_i}{n} \bar{X}_{ni}^{(u)} - m_u \right| &\leq \left| \sum_{J \in J_n^*} t_J^{(u)} \mu(J) - m_u \right| + \left| \sum_{i=1}^{c_n} \frac{n_i}{n} \bar{X}_{ni}^{(u)} - \sum_{J \in J_n^*} t_J^{(u)} \mu(J) \right| \\ &\leq \left| \sum_{J \in J_n^*} t_J^{(u)} \mu(J) - m^{(u)} \right| + \left| \sum_{i=1}^{c_n} \left(\frac{n_i}{n} - \mu(I_{ni}) \right) \bar{X}_{ni}^{(u)} \right| + \left| \sum_{J \in \tilde{J}_n} t_J^{(u)} \mu(J) \right| \\ &\triangleq L_1 + L_2 + L_3. \end{aligned} \quad (21)$$

Since $E\|\bar{X}\| < \infty$, $L_1 \rightarrow 0$ as $n \rightarrow \infty$. As for L_3 , when event E_n occurs, we have $\tilde{J}_n \subset F_0^c \cup \{a_{N+1}, \dots, a_{N+2}\}$ and

$$\limsup_{n \rightarrow \infty} \left| \sum_{J \in \tilde{J}_n} t_J^{(u)} \mu(J) \right| \leq E(|\bar{X}^{(u)}| | I(F_0^c \cup \{a_{N+1}, a_{N+2}, \dots\})). \quad (22)$$

By definition of N and F_0 , we have $P(\bar{X} \in F_0^c \cup \{a_{N+1}, a_{N+2}, \dots\}) < 3\eta$. Letting $\eta > 0$ small enough, the right hand side of (22) can be made small enough.

This, combining with (18), shows that

$$\lim_{n \rightarrow \infty} L_3 = 0, \quad \text{a.s.} \quad (23)$$

Finally we turn to L_2 . When event F_n occurs, we have $\mu(I_{ni}) \geq n^{-1/3+\epsilon_1/2}$.

Also, by (17) we have

$$\begin{aligned} P(|\#(I_{ni})/n - \mu(I_{ni})| \geq \epsilon \mu(I_{ni}) \text{ for at least one } i \text{ such that} \\ \mu(I_{ni}) \geq n^{-1/3+\epsilon_1/2}) \leq n \cdot 2 \exp(-c n n^{-1/3+\epsilon_1/2}) = o(n^{-2}). \end{aligned} \quad (24)$$

This, combining with (19) (choosing $k = 2$ in (19)), shows that with probability one $|n_i/n - \mu(I_{ni})| \leq \epsilon \mu(I_{nk})$, $i = 1, \dots, c_n$, for n sufficiently large. In this event we have

$$L_2 \leq \epsilon \sum_{i=1}^{c_n} |\bar{X}_{ni}^{(u)}| \mu(I_{ni}) \leq \epsilon \sum_{J \in J_n^*} |t_J^{(u)}| \mu(J) \rightarrow \epsilon E|\bar{X}^{(u)}|.$$

Therefore

$$\lim_{n \rightarrow \infty} L_2 = 0, \quad \text{a.s.}$$

and we have proved that

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{c_n} \frac{n_i}{n} \bar{X}_{ni}^{(u)} = m_u, \quad u = 1, \dots, d, \quad \text{a.s.} \quad (25)$$

Similar arguments shows that

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{c_n} \frac{n_i}{n} \bar{X}_{ni}^{(u)} \bar{X}_{ni}^{(v)} = E(\bar{X}^{(u)} \bar{X}^{(v)}), \quad u, v = 1, \dots, d, \quad \text{a.s.} \quad (26)$$

Now (14) follows from (25), (26).

According to an important theorem in the theory of strong approximation proved by Komlós and others (1975), one can find a random variable

$Z_{ni} \sim N(0, \sigma^2/n_i)$, Z_{ni} , ..., Z_{nc_n} independent, and

$$P(|e_{ni} - Z_{ni}| \geq (\log n_i)^2/n_i) \leq K \exp(-\lambda(\log n_i)^2), \quad i = 1, \dots, c_n \quad (27)$$

where $K > 0$, $\lambda > 0$ are constants not depending on n , i . Put

$$V_{ni} = e_{ni} - Z_{ni}, \quad i = 1, \dots, c_n$$

$$\tilde{V}_{ni} = V_{ni} - \frac{1}{N_n} \sum_{i=1}^{c_n} n_i V_{ni}, \quad \tilde{Z}_{ni} = Z_{ni} - \frac{1}{N_n} \sum_{i=1}^{c_n} n_i Z_{ni}, \quad i = 1, \dots, c_n$$

$$V_{(n)} = (\tilde{V}_{ni}, \dots, \tilde{V}_{nc_n})', \quad Z_{(n)} = (\tilde{Z}_{ni}, \dots, \tilde{Z}_{nc_n})'$$

$$\xi_n = Z_{(n)}' W_{(n)} Z_{(n)} - Z_{(n)}' W_{(n)} \bar{X}_{(n)} (\bar{X}_{(n)}' W_{(n)} \bar{X}_{(n)})^{-1} \bar{X}_{(n)}' W_{(n)} Z_{(n)} \quad (28)$$

Then RSS_n defined by (19) can be written as

$$\begin{aligned} RSS_n &= \xi_n + V'_{(n)} W_{(n)} V_{(n)} + 2V'_{(n)} W_{(n)} Z_{(n)} \\ &\quad - V'_{(n)} W_{(n)} \bar{X}_{(n)} (\bar{X}'_{(n)} W_{(n)} \bar{X}_{(n)})^{-1} \bar{X}'_{(n)} W_{(n)} V_{(n)} \\ &\quad - 2V'_{(n)} W_{(n)} \bar{X}_{(n)} (\bar{X}'_{(n)} W_{(n)} \bar{X}_{(n)})^{-1} \bar{X}'_{(n)} W_{(n)} Z_{(n)} \\ &\triangleq \xi_n + Q_1 + 2Q_2 - Q_3 - 2Q_4. \end{aligned} \quad (29)$$

From the well-known normal theory and the assumptions made on model (1), it follows that the conditional distribution of ξ_n given $\{\bar{X}_1, \dots, \bar{X}_n\}$ is $\chi^2_{c_n-d}$ (so the unconditional distribution of ξ_n is also $\chi^2_{c_n-d}$). Now we proceed to estimate $Q_1 - Q_4$. First take Q_1 . Considering (6), we see that the event

$$T_n = \bigcup_{i=1}^{c_n} \{|e_{ni} - Z_{ni}| \geq (\log n_i)^2/n_i\}$$

has a probability $O(n^{-2})$. Hence, with probability one, we can assert that

$$|v_{ni}| \leq (\log n_i)^2/n_i, \quad i = 1, \dots, c_n \quad (30)$$

for n sufficiently large. Now, in case (30) is true.

$$Q_1 = \sum_{i=1}^{c_n} n_i \tilde{v}_{ni} \leq \sum_{i=1}^{c_n} n_i v_{ni} \leq \sum_{i=1}^{c_n} (\log n_i)^4/n_i.$$

Considering (5) and (6), we have (again in case (30) is true)

$$Q_1 \leq n^{1/3-\epsilon_1} (\log n)^4 n^{-2/3-\epsilon_1} = n^{-1/3-2\epsilon_1} (\log n)^4 \quad (31)$$

which proves that

$$\lim_{n \rightarrow \infty} Q_1 = 0, \quad \text{a.s.} \quad (32)$$

Next come to Q_2 . By Schwartz inequality,

$$Q_2^2 \leq Q_1 \cdot Z'_{(n)} W_{(n)} Z_{(n)} \leq Q_1 \sum_{i=1}^{c_n} n_i Z_{ni}^2 \quad (33)$$

But $\sum_{i=1}^{c_n} n_i Z_{ni}^2 \sim x_{c_n}^2$. Hence $\sum_{i=1}^{c_n} n_i Z_{ni}^2 / c_n = O_p(1)$, as $n \rightarrow \infty$. From this, (5), and (31), we see that

$$Q_2 \xrightarrow{P^*} 0, \text{ as } n \rightarrow \infty. \quad (34)$$

In order to deal with Q_3 , we shall in the following use the symbol k_n to denote a quantity depending on $\bar{X}_1, \dots, \bar{X}_n$, with the property that $\sup_n k_n < \infty$ with probability one. k_n may assume different values in each of its appearance. We also use the symbol I to denote an identity matrix with appropriate order. Under these conventions, we have by (14) that

$$(\bar{X}'_{(n)} W_{(n)} \bar{X}_{(n)})^{-1} \leq k_n I/n \quad (35)$$

for n sufficiently large. Therefore,

$$\begin{aligned} Q_3 &\leq V'_{(n)} W_{(n)}^{1/2} \bar{X}_{(n)}^{1/2} (k_n I/n) \bar{X}'_{(n)} W_{(n)}^{1/2} V_{(n)} \\ &\leq k_n V'_{(n)} W_{(n)}^{1/2} \left(\sum_{i=1}^{c_n} \frac{n_i}{n} \frac{\bar{X}_i}{\bar{X}_{ni} - \bar{X}_{ni}} \right) W_{(n)}^{1/2} V_{(n)}. \end{aligned}$$

By this, and using again (14), we have

$$Q_3 \leq k_n V'_{(n)} W_{(n)}^{1/2} W_{(n)}^{1/2} V_{(n)} = k_n Q_1 \quad (36)$$

with probability one for n sufficiently large. By (32), we get

$$\lim_{n \rightarrow \infty} Q_3 = 0, \text{ a.s.} \quad (37)$$

Finally we come to Q_4 . By Schwartz inequality,

$$Q_4^2 \leq Q_3 Z'_{(n)} W_{(n)} \bar{X}_{(n)} (\bar{X}'_{(n)} W_{(n)} \bar{X}_{(n)})^{-1} \bar{X}'_{(n)} W_{(n)} Z_{(n)}. \quad (38)$$

Arguing as in the case of Q_3 , we have

$$Z'_{(n)} W_{(n)} \bar{X}_{(n)} (\bar{X}'_{(n)} W_{(n)} \bar{X}_{(n)})^{-1} \bar{X}'_{(n)} W_{(n)} Z_{(n)} \leq k_n Z'_{(n)} W_{(n)} Z_{(n)} \quad (39)$$

with probability one for n sufficiently large. From (36), (38), (39), we have (again with probability one for n sufficiently large)

$$Q_4^2 \leq k_n Q_1 Z'_{(n)} W_{(n)} Z_{(n)}. \quad (40)$$

In the process of dealing with Q_2 , we have shown that the right hand side of (40) tends to zero in probability as $n \rightarrow \infty$. Hence, with probability one,

$$Q_4 \xrightarrow{P^*} 0, \quad \text{as } n \rightarrow \infty. \quad (41)$$

From (29), (32), (34), (37), (41), and the observation made on ξ_n , we finally reach (13). Theorem 1 is proved.

5. ESTIMATION OF THE VARIANCE OF ERROR

In this section we shall give an estimate of the variance σ^2 . For our purpose, we need an estimator σ_n^2 which tends to σ^2 with a rate as $O(c_n^{-1/2})$.

We shall consider separately the two cases that $1^\circ, d \leq 3, 2^\circ, d > 3$. For the first case, we choose ϵ_1, ϵ_2 satisfying (2) and

$$\epsilon_2 < (9\epsilon_1 + 1)/12 \quad (42)$$

and define

$$\sigma_n^2 = \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} (Y_{ni}(j) - Y_{ni})^2 / (N_n - c_n) \quad (43)$$

For the second case, we choose ϵ_1, ϵ_3 , such that

$$0 < \epsilon_1 < 1/3, \quad 2(1 - \epsilon_3)/d > \frac{1}{6} - \frac{\epsilon_1}{2}. \quad (44)$$

Define $q_n = n^{-(1-\epsilon_3)/d}$. Decompose R^d into a set \tilde{J}_n of supercubes having the form (3), but λ_n in (3) is changed to q_n . For each $J \in \tilde{J}_n$, write the elements in $\{\bar{X}_1, \dots, \bar{X}_n\} \cap J$ as $\{\bar{X}_{nJ}(1), \dots, \bar{X}_{nJ}(n_J)\}$. The observations of Y corresponding to these \bar{X} 's is written as $\{Y_{nJ}(1), \dots, Y_{nJ}(n_J)\}$.

Finally we define

$$\sigma_n^2 = \sum_{J \in \tilde{J}_n} \sum_{j=1}^{n_J} (Y_{nJ}(j) - Y_{nJ})^2 / \sum_{J \in \tilde{J}_n} (n_J - 1) \quad (45)$$

where $Y_{nJ} = \sum_{j=1}^{n_J} Y_{nJ}(j) / n_J$.

THEOREM 2. Under model (1) and suppose that H_0 is true, then with probability 1

$$\sqrt{c_n}(\sigma_n^2 - \sigma^2) \xrightarrow{P^*} 0, \quad \text{as } n \rightarrow \infty. \quad (46)$$

Proof. We shall only give the details for the case $d \leq 3$, as the case $d > 3$ can be dealt with in an entirely similar way. Put $b_{ni}(j) = (\bar{X}_{ni}(j) - \bar{X}_{ni})'\beta$, then $|b_{ni}(j)| \leq \sqrt{d}\|\beta\|_{\mathcal{L}_n}$. From (18), we have

$$\begin{aligned} \sum_{j=1}^{n_i} (Y_{ni}(j) - \bar{Y}_{ni})^2 &= \sum_{j=1}^{n_i} (e_{ni}(j) - \bar{e}_{ni})^2 \\ &+ \sum_{j=1}^{n_i} b_{ni}^2(j) + 2 \sum_{j=1}^{n_i} b_{ni}(j)(e_{ni}(j) - \bar{e}_{ni}). \end{aligned} \quad (47)$$

Put $N'_n = N_n - C_n$. In view of (5), in order to prove (46), it is enough to verify the following assertions (all with probability one).

$$1^\circ. \quad n^{1/6-\epsilon_1/2} T_n / N'_n \xrightarrow{P^*} 0, \quad \text{where } T_n = \sum_{i=1}^{C_n} \sum_{j=1}^{n_i} (e_{ni}^2(j) - \sigma^2).$$

$$2^\circ. \quad n^{1/6-\epsilon_1/2} \sum_{i=1}^{C_n} n_i e_{ni}^2 / N'_n \xrightarrow{P^*} 0.$$

$$3^\circ. \quad n^{1/6-\epsilon_1/2} \sum_{i=1}^{C_n} \sum_{j=1}^{n_i} b_{ni}^2(j) / N'_n \xrightarrow{P^*} 0.$$

$$4^\circ. \quad S_n \xrightarrow{P^*} 0, \quad \text{where } S_n = n^{1/6-\epsilon_1/2} \sum_{i=1}^{C_n} \sum_{j=1}^{n_i} b_{ni}(j)(e_{ni}(j) - \bar{e}_{ni}) / N'_n.$$

Under model (1), T_n is a sum of N_n iid. random variables with zero mean and finite variance v . By (6), we have $N'_n \geq n^{2/3+\epsilon_1} - 1 \geq \frac{1}{2} n^{2/3+\epsilon_1}$. Hence (with probability one)

$$P^*(|n^{1/6-\epsilon_1/2} T_n / N'_n| \geq \epsilon) \leq P^*(|T_n| \geq \frac{\epsilon}{2} n^{1/2+3\epsilon_1/2}) \leq \frac{4}{\epsilon^2} v n^{-3\epsilon_1} \rightarrow 0.$$

This proves 1° . For 2° , notice that $E(n_i e_{ni}^2) = \sigma^2$, hence

$$E^*(n^{1/6-\epsilon_1/2} \sum_{i=1}^{C_n} n_i e_{ni}^2 / N'_n) \leq n^{1/6-\epsilon_1/2} C_n / N'_n \leq 2n^{-5\epsilon_1/2} \rightarrow 0$$

This proves 1°. For 2°, notice that $E(n_i e_{ni}^2) = \sigma^2$, hence

$$E^*(n^{1/6-\epsilon_1/2} \sum_{i=1}^{c_n} n_i e_{ni}^2 / N'_n) \leq n^{1/6-\epsilon_1/2} C_n / N'_n \leq 2n^{-5\epsilon_1/12} \rightarrow 0$$

and 2° follows. 3° is proved by noticing that $N'_n / N_n \rightarrow 1$, and

$$\begin{aligned} n^{1/6-\epsilon_1/2} \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}^2(j) / N'_n &\leq \frac{N_n}{N'_n} d \|\beta\|^2 n^{1/6-\epsilon_1/2} \\ &= \frac{N_n}{N'_n} d \|\beta\|^2 n^{-\frac{1}{d}(2/3-2\epsilon_2)} + (1/6-\epsilon_1/2) \end{aligned}$$

under the condition $0 < \epsilon_1 < 1/3$, $0 < \epsilon_2 < (9\epsilon_1 + 1)/12$ and $d \leq 3$, we have $-(2/3 - 2\epsilon_2)/d + (1/6 - \epsilon_1/2) < 0$. Therefore, the right hand side tends to zero as $n \rightarrow \infty$.

Finally we come to 4°. Since the sum $\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} d_{ni}(j)(e_{ni}(j) - e_{ni})$ can be written in a form $\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} d_{ni}(j)e_{ni}(j)$ with $|d_{ni}(j)| \leq |b_{ni}(j)|$, it follows that

$$\text{Var}^*(S_n) \leq n^{1/3-\epsilon_1} \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}^2(j) \sigma^2 / N_n'^2.$$

It is easy to see that the right hand side is less than a constant g_n which does not depend on $\bar{X}_1, \dots, \bar{X}_n$, and $g_n \rightarrow 0$ as $n \rightarrow \infty$. Also, $E^*(S_n) = 0$. Therefore,

$$\begin{aligned} E^*S_n^2 &= \text{Var}^*S_n + (E^*(S_n))^2 \\ &\leq g_n \rightarrow 0, \end{aligned}$$

from which 4° follows, and Theorem 2 is proved.

A more tedious argument allows us to prove that

$$\lim_{n \rightarrow \infty} \sqrt{C_n}(\sigma_n^2 - \sigma^2) = 0, \quad \text{a.s.}$$

6. A LARGE SAMPLE DETERMINATION OF C IN TEST (12)

We are now in a position to prove the following theorem, which gives a large sample determination of C in test (12).

THEOREM 3. Under model (1) and suppose that the null hypothesis H_0 is true, the distribution of the statistic

$$T_n^* \triangleq \sqrt{2RSS_n / \frac{2}{n}} - \sqrt{2(c_n - d)}$$

converges to the standard normal $N(0,1)$ as $n \rightarrow \infty$.

Proof. From Theorem 1 it follows that the conditional distribution of $\tilde{T}_n \triangleq \sqrt{2RSS_n / \sigma^2} - \sqrt{2(c_n - d)}$ given $\{\bar{X}_1, \bar{X}_2, \dots\}$ tends to $N(0,1)$, with probability one. Since $|\sigma_n - \sigma|^2 \leq |\sigma_n^2 - \sigma^2|$, we have

$$|\tilde{T}_n - T_n^*| = \sqrt{2RSS_n} \frac{|\sigma - \sigma_n|}{\sigma\sigma_n} \leq \frac{\sqrt{2RSS_n} |\sigma_n^2 - \sigma^2|}{\sigma\sigma_n}. \quad (48)$$

By Theorem 1, with probability one $RSS_n / c_n = O_p^*(1)$. Therefore, from (48) and Theorem 2, we have $\tilde{T}_n - T_n^* \xrightarrow{P^*} 0$, with probability one. Hence we have proved that with probability one, the conditional distribution of T_n^* given $\{\bar{X}_1, \bar{X}_2, \dots\}$ tends to $N(0,1)$, as $n \rightarrow \infty$, and the same holds true for the unconditional distribution of T_n^* . Theorem 3 is proved.

From Theorem 3 it follows that if we choose

$$C = \{\sqrt{2(c_n - d)} + u_{\alpha_0}\}^2 / 2 \quad (49)$$

in (12), where u_{α_0} is defined by $\int_{-\infty}^{u_{\alpha_0}} e^{-x^2/2} dx / \sqrt{2\pi} = 1 - \alpha_0$, then the test is asymptotically similar with size α_0 .

7. CONSISTENCY OF THE TEST

Denote by $Q_n(F)$ the power of the test defined by (12) and (49), where F is the joint distribution of (\bar{X}, Y) . The following theorem shows that this test is consistent over a wide class of alternatives.

THEOREM 4. Suppose that the following conditions are satisfied:

- 1°. $E\|\bar{X}\|^2 < \infty$, $\text{COV}(\bar{X}) > 0$, $EY^2 < \infty$.
- 2°. The distribution of \bar{X} has no singular component.
- 3°. The closure \bar{A}_1 of the set A of all atoms of \bar{X} possesses Lebesgue measure zero.
- 4°. The closure \bar{A}_2 of all discontinuity points of $m(x) = E(Y|\bar{X} = x)$ possesses Lebesgue measure zero.
- 5°. $P(Y < y|\bar{X} = x)$ depends only on $y - m(x)$.
- 6°. There exists no linear function $\alpha + \beta'x$ such that

$$P(m(\bar{X}) = \alpha + \beta'\bar{X}) = 1.$$

Then, denoting the distribution of (\bar{X}, Y) by F , we have

$$\lim_{n \rightarrow \infty} Q_n(F) = 1. \quad (50)$$

Proof. First choose $d+2$ points t_1, \dots, t_{d+2} , such that the points

$$(t'_1, m(t_1)), \dots, (t'_{d+2}, m(t_{d+2})) \quad (51)$$

do not lie on a common hyperplane in R^{d+1} . If \bar{X} possesses no absolute continuous component, then, by conditions 1° and 6°, we can choose such points t_1, \dots, t_{d+2} from the atom set A . If \bar{X} has an absolute continuous component with density f , then, according to conditions 1°, 3°, 4°, 6°, we can choose t_1, \dots, t_{d+2} from the set $R^d - (\bar{A}_1 \cup \bar{A}_2)$. With the additional

property that there exists two constants $h_1 > 0$, $h_2 > 0$, such that for any $I = \{(x^{(1)}, \dots, x^{(d)}) : b_i \leq x^{(i)} < b_i + h, i = 1, \dots, d\}$ with $h \in (0, h_1)$, we have

$$\int_I f dx \geq h_2 h^d. \quad (52)$$

For large n , we can find $d+2$ distinct supercubes $J_{n1}, \dots, J_{n,d+2}$ in J_n^* (see (3)), such that $t_i \in J_{ni}$, $i = 1, \dots, d+2$. Denote the elements of $\{\bar{X}_1, \dots, \bar{X}_n\} \cap J_{ni}$ by $\bar{X}_{ni}(j)$, $j = 1, \dots, n_i$. In the same way as we did in Section 2, we define $Y_{ni}(j)$, \bar{X}_{ni} , Y_{ni} . Fit a hyperplane

$$y = \alpha_n + \beta_n' x$$

to the $d+2$ points (\bar{X}_{ni}, Y_{ni}) , $i = 1, \dots, d+2$, by the weighted least squares method with weights n_1, \dots, n_{d+2} . The weighted residual sum of squares is

$$R_n = \sum_{i=1}^{d+2} n_i (Y_{ni} - \alpha_n - \beta_n' \bar{X}_{ni})^2.$$

For definiteness, we assume that the points t_1, \dots, t_{d+2} are chosen from the set $R^d - (\bar{A}_1 \cup \bar{A}_2)$. In this case, from (52) and the fact that $\ell_n = -(1/3 - \epsilon_2)^d$, With probability one we have

$$n_i \geq h_2 n^{2/3 + \epsilon_2} / 2, \quad i = 1, \dots, d+2 \quad (53)$$

for n sufficiently large.

Given $\delta_1 > 0$, we can find n_0 such that for $n \geq n_0$

$$\sup\{m(x) : x \in J_{ni}\} - \inf\{m(x) : x \in J_{ni}\} < \delta_1 / 2, \quad i = 1, \dots, d+2. \quad (54)$$

Use again $P^* = P^*(\bar{X}_1, \bar{X}_2, \dots)$ the conditional probability distribution given $\bar{X}_1, \bar{X}_2, \dots$. By (53), (54), it is easy to see that

$$P^*(|Y_{ni} - m(t_i)| \geq \delta_1) \leq \delta_1^{-2} h_2^{-1} \sigma_n^{-2/3 - \epsilon_2}, \quad i = 1, \dots, d+2 \quad (55)$$

with probability one for n sufficiently large.

Since the points (51) do not lie simultaneously on any hyperplane, and since $\bar{X}_{ni} \rightarrow t_i$, $i = 1, \dots, d+2$, as $n \rightarrow \infty$, it follows that there exists constant $\delta_2 > 0$ not depending on n and (\bar{X}_i, Y_i) , $i = 1, 2, \dots$, such that

$$\sup\{|m(t_i) - \alpha_n - \beta_n' \bar{X}_{ni}| : i = 1, \dots, d+2\} \geq \delta_2$$

for n sufficiently large. Choose $\delta_1 = \delta_2/2$ in (54). Then if the inequalities $|Y_{ni} - m(t_i)| < \delta_1$ all hold true, we shall have

$$R_n \geq \delta_1^2 \min\{n_1, \dots, n_{d+2}\} \quad (56)$$

for n sufficiently large. Summing up all the arguments above, we reach the following conclusion:

$$\lim_{n \rightarrow \infty} P(R_n > n^{2/3}) = 1. \quad (57)$$

Obviously, we have $RSS_n \geq R_n$. Hence by (57)

$$\lim_{n \rightarrow \infty} P(RSS_n > n^{2/3}) = 1. \quad (58)$$

Now consider σ_n^2 defined by (45). The equality (47) still holds if we redefine $b_{ni}(j) = m(\bar{X}_{ni}(j)) - m_{ni}$, with $m_{ni} = \sum_{j=1}^{n_i} m(\bar{X}_{ni}(j))/n_i$. We have

$$\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}^2(j)/N_n \leq \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} m^2(\bar{X}_{ni}(j))/N_n \leq \sum_{i=1}^n m^2(\bar{X}_i)/N_n. \quad (59)$$

It can be shown that there exists constant $q > 0$, such that with probability one we have

$$N_n \geq qn \quad (60)$$

for n sufficiently large (see the end of this section). From (59), (60), using the strong law of large numbers, we get

$$\limsup_{n \rightarrow \infty} \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}^2(j)/N_n \leq q^{-1} E(m^2(\bar{X})) \leq q^{-1} E(Y^2) < \infty, \text{ a.s.} \quad (61)$$

Employing the method of handling S_n in the proof of Theorem 2, we get

$$\begin{aligned} P^*(\left| \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}(j)(e_{ni}(j) - e_{ni}) \right|/N_n \geq \epsilon) &\leq \epsilon^{-2} N_n^{-2} \sigma^2 \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}^2(j) \\ &\leq \epsilon^{-2} N_n^{-2} \sigma^2 \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} m^2(\bar{X}_{ni}(j)) \leq \epsilon^{-2} N_n^{-2} \sigma^2 \sum_{i=1}^n m^2(\bar{X}_i) \rightarrow 0, \text{ a.s.} \end{aligned} \quad (62)$$

Finally, we have

$$\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} (e_{ni}(j) - e_{ni})^2 / N_n \xrightarrow{P^*} 0. \quad (63)$$

Combining (61)-(63), we see that there exists constant $M < \infty$, such that

$$\lim_{n \rightarrow \infty} P(\sigma_n^2 < M) = 1. \quad (64)$$

From (58), (64), it follows that

$$\lim_{n \rightarrow \infty} P(RSS_n / \sigma_n^2 > n^{2/3} M^{-1}) = 1. \quad (65)$$

But by (5), we have

$$C < (\sqrt{2(c_n - d)} + u_{\alpha_0})^2 < 8c_n < 8n^{1/3} < M^{-1} n^{2/3}$$

for n sufficiently large. This means

$$\lim_{n \rightarrow \infty} P(RSS_n / \sigma_n^2 > C) = 1$$

which is no other than (50), and the theorem is proved.

Proof of (60). If \bar{X} has at least one atom, then clearly (60) holds.

Otherwise we choose a sufficiently large such that

$$\int_I f dx > 1/2, \quad I = \{(x^{(1)}, \dots, x^{(d)}) : |x^{(i)}| \leq a, i = 1, \dots, d\}.$$

I contains $(2a)^d n^{1/3-\epsilon_2}$ elements of J_n^* . Put

$$H_n = \{J : J \in J_n^*, J \subset I, \mu(J) \geq 4^{-1}(2a)^{-d} n^{-(1/3-\epsilon_2)}\}$$

Evidently we have

$$\sum_{J \in H_n} \mu(J) \geq 1/4. \quad (66)$$

Using (16), and noticing that H_n has at most $(2a)^d n^{1/3-\epsilon_2}$ members, we obtain

$$\begin{aligned} P(|\#(J)/n - \mu(J)| \geq 8^{-1}(2a)^{-d} n^{-(1/3-\epsilon_2)}, \text{ for at least one } J \in H_n) \\ \leq (2a)^d n^{1/3-\epsilon_2} 2 \exp(-cn^{1/3+2\epsilon_2}) = O(n^{-2}) \end{aligned} \quad (67)$$

where $c > 0$ does not depend on n . From (67) we see that, with probability one, we have

$$|\#(J)/n - \mu(J)| < 8^{-1}(2a)^{-d} n^{-(1/3-\epsilon_2)}, \text{ for all } J \in H_n \quad (68)$$

for n sufficiently large. But when (68) holds we shall have

$$\begin{aligned} N_n &\geq \sum_{J \in H_n} \#(J) \geq n \sum_{J \in H_n} [\mu(J) - 8^{-1}(2a)^{-d} n^{-(1/3-\epsilon_2)}] \\ &\geq n \sum_{J \in H_n} \frac{1}{2} \mu(J) \geq n/8 \end{aligned}$$

on account of the fact that $\epsilon_2 > \epsilon_1$ and (66). This shows (60) holds with $q = 1/8$.

8. ASYMPTOTIC POWER OF THE TEST

Suppose that $(\bar{X}, Y^{(1)}), (\bar{X}, Y^{(2)}), \dots$ is a sequence of alternatives, where $m^{(n)}(x) \triangleq E(Y^{(n)} | \bar{X} = x)$ approaches some linear function $\alpha + \beta'x$ as the sample size n increases. Specifically we assume that

$$m^{(n)}(x) = \alpha + \beta'x + m_0(x)/g_n \quad (69)$$

where $g_n \rightarrow \infty$ as $n \rightarrow \infty$. We are interested in finding such g_n for which $Q_n(F^{(n)})$ tends to some limit greater than the size α_0 and smaller than one ($F^{(n)}$ is the distribution of $(\bar{X}, Y^{(n)})$).

THEOREM 5. Suppose that $(\bar{X}_1, Y_1), \dots, (\bar{X}_n, Y_n)$ are iid. samples of $(\bar{X}, Y^{(n)})$, where $Y_i = m^{(n)}(\bar{X}_i) + e_i$, $i = 1, \dots, n$, $m^{(n)}(x)$ is defined by (69) and \bar{X}_i, e_i , $i = 1, \dots, n$ satisfy the conditions specified in model (1). Also, assume that $EM_0^2(\bar{X}) < \infty$, and

$$\lim_{n \rightarrow \infty} g_n / (\sqrt{n} c_n^{-1/4}) = h^{-1} \in (0, \infty). \quad (70)$$

Define

$$\ell = u_{\alpha_0} - \frac{h^2}{\sqrt{2}} \min_{\alpha, \beta} E[m_0(\bar{X}) - \alpha - \beta'\bar{X}]^2. \quad (71)$$

Then we have

$$\lim_{n \rightarrow \infty} Q_n(F^{(n)}) = \frac{1}{\sqrt{2\pi}} \int_{\ell}^{\infty} e^{-t^2/2} dt. \quad (72)$$

Proof. As the detailed proof is tedious, we give only a sketch of the main points of the proof.

1. Since the linear hypothesis H_0 is not assumed to be true, the residual sum of squares RSS_n (9) cannot be reduced to (11). Under this circumstance, instead of (27), we have to use a strong approximation to Y_{ni} :

$P(|Y_{ni} - Z_{ni}^*| \geq (\log n_i)^2/n_i) \leq K \exp(-\lambda(\log n_i)^2)$, $i = 1, \dots, c_n$
 where $Z_{ni}^* \sim N(u_{ni}, \sigma^2/n_i)$, with

$$u_{ni} = \alpha + \beta' \bar{X}_{ni} + \sum_{j=1}^{c_n} m_0(\bar{X}_{ni}(j))/(n_i g_n).$$

2. Put $Z_{ni} = Z_{ni}^* - \alpha - \beta' \bar{X}_{ni}$. It is easy to verify that the expression (29) still holds true with Z_{ni} so defined.

3. We have to verify that (32), (34), (37) and (41) still holds true in the present case. For Q_1 , nothing has been changed and (31) is true. For Q_2 , notice that $Z_{(n)}' W_{(n)} Z_{(n)} / \sigma^2$ obeys a non-central χ^2 distribution with degree of freedom c_n and non-central parameter

$$\tilde{\delta}_n^2 = g_n^{-2} \sum_{i=1}^{c_n} \left(\sum_{j=1}^{n_i} m_0(\bar{X}_{ni}(j))/n_i \right)^2.$$

It is easily seen that as $n \rightarrow \infty$, $\tilde{\delta}_n^2 = ng_n^{-2} [Em_0^2(\bar{X}) + o(1)]$. Since we have $ng_n^{-2} = O(\sqrt{c_n})$ in view of (70), by (5), we have

$$E(Z_{(n)}' W_{(n)} Z_{(n)}) = O(c_n) = O(n^{1/6 - \epsilon_1/2}) \quad (73)$$

and (34) follows from (31) and (73). Q_3 and Q_4 can be dealt with in the same manner.

4. Given $\bar{X}_1, \bar{X}_2, \dots$, the conditional distribution of ε_n / σ^2 is $\chi_{c_n, \delta_n^2}^2$, where

$$\delta_n^2 = ng_n^{-2} \left\{ \sum_{i=1}^{c_n} \frac{n_i}{n} v_{ni}^2 - \left(\sum_{i=1}^{c_n} \frac{n_i}{n} v_{ni} \bar{X}_{ni} \right) \left(\sum_{i=1}^{c_n} \frac{n_i}{n} (\bar{X}_{ni} - \bar{\bar{X}}_n) (\bar{X}_{ni} - \bar{\bar{X}}_n)' \right) \left(\sum_{i=1}^{c_n} \frac{n_i}{2} v_{ni} \bar{X}_{ni} \right) \right\}$$

where

$$v_{ni} = \sum_{j=1}^{n_i} m_0(\bar{X}_{ni}(j))/n_i - m_0, \quad m_0 = Em_0(\bar{X}).$$

From (14) and (70), it is easily seen

$$\begin{aligned}
\delta_n^2 &= \sqrt{c_n} \{ \text{Var}(m_0(\bar{X})) - E[(m_0(\bar{X}) - m_0)\bar{X}'] V^{-1} E[(m_0(\bar{X}) - m_0)\bar{X}] + o(1) \} h^2 \\
&= \sqrt{c_n} \{ \min_{\alpha, \beta} E(m_0(\bar{X}) - \alpha - \beta' \bar{X})^2 + o(1) \} h^2.
\end{aligned} \tag{74}$$

Given $\{\bar{X}_1, \bar{X}_2, \dots\}$, the conditional distribution of ξ_n/σ^2 is the same as the sum of iid. variables $(\eta_i + \delta_n/\sqrt{c_n})^2$, $i = 1, \dots, c_n$, where $\eta_1, \dots, \eta_{c_n}$ are iid., $\eta_1 \sim N(0,1)$. The central limit theorem can be applied to this case, which gives

$$\frac{\xi_n/\sigma^2 - (c_n + \delta_n^2)}{\sqrt{2c_n + 4\delta_n^2}} \rightarrow N(0,1). \tag{75}$$

5. Define σ_n^2 as in (45). We have to verify that (46) still holds true in the present case. Starting from (47), in which $b_{ni}(j)$ is redefined as

$$\begin{aligned}
b_{ni}(j) &= (\bar{X}_{ni}(j) - \bar{X}_{ni})' \beta + g_n^{-1} u_{ni}(j) \\
u_{ni}(j) &= m_0(\bar{X}_{ni}(j)) - \sum_{j=1}^{n_i} m_0(\bar{X}_{ni}(j))/n_i
\end{aligned}$$

it is easily seen that the first and third term in the right hand side of (47) can be handled in the same way as before, and it suffices to verify that

$$\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} b_{ni}^2(j)/n = o(\sqrt{c_n}). \tag{76}$$

Considering (46), (75) reduces to

$$g_n^{-2} \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} u_{ni}^2(j)/n = o(\sqrt{c_n}) \tag{77}$$

Since

$$\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} u_{ni}^2(j)/n \leq \sum_{i=1}^{c_n} \sum_{j=1}^{n_i} m_0^2(\bar{X}_{ni}(j))/n \leq \sum_{i=1}^n m_0^2(\bar{X}_i)/n \rightarrow E(m_0^2(\bar{X})), \text{ a.s.}$$

Further, by (5) and (70), we have

$$g_n^{-2} \leq \frac{1}{2} h^2 n^{-1} \sqrt{c_n} \leq n^{-5/6}$$

for n sufficiently large. This proves (77).

6. From (45) and (75), we have

$$\frac{\varepsilon_n/\sigma_n^2 - (c_n + \delta_n^2)}{\sqrt{2c_n + 4\delta_n^2}} \rightarrow N(0,1)$$

as $n \rightarrow \infty$, under the conditional distribution given $\{\bar{X}_1, \bar{X}_2, \dots\}$. From (49), (74), we have

$$\lim_{n \rightarrow \infty} \frac{(\sqrt{2(c_n - d)^2 + u_{\alpha_0}^2}/2 - (c_n + \delta_n^2))}{\sqrt{2c_n + 4\delta_n^2}} = \lambda.$$

This leads to (72), concluding the proof of the theorem.

The statistical meaning of the quantity λ is clear, since

$\min_{\alpha, \beta} E[m_0(\bar{X}) - \alpha - \beta'\bar{X}]^2$ measures the distance of $m_0(\bar{X})$ to linearity in the MSE sense. As this distance increases, λ decreases, and the asymptotic power of the test (right hand side of (72)) increases.

The theorem indicates that, roughly speaking, our test has a discrimination power for those Y , for which the distance (in the sense of MSE) of $E(Y|\bar{X} = x)$ to the linear function space is not smaller than $O(n^{-1/2} c_n^{1/4})$. Under quite general conditions, we can prove that $c_n \sim Cn^{1/3-\varepsilon_2}$, and $O(n^{-1/2} c_n^{1/4}) = O(n^{-5/12-\varepsilon_2/4})$. In principle, ε_2 can be chosen arbitrarily near $1/3$. So the order for possible discrimination can be made arbitrarily

near $O(n^{-1/2})$, but this order cannot be reached, unless \bar{X} is purely atomic with a finite number of atoms.

REFERENCES

- [1] BENNETT, G. (1962). Probability Inequalities for the Sums of Independent Random Variables. *J. Amer. Statist. Assoc.*, 57, 33-45.
- [2] CSORGO, M. and REVESZ, P. (1981). Strong Approximations in Probability and Statistics, *Akademiai Kiadó*.
- [3] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58, 11-30.
- [4] KOMLOS, J., MAJOR, P., TUSNADY, G. (1975). An Approximation of Partial Sums of Independent R.V.'s and the Sample DF I. *Z. Wahrsh. Verw. Gebiete* 32, 111-131.

END

12-87

DTIC